DOCUMENT RESUME

ED 270 488 TM 860 371

AUTHOR Braungart-Bloom, Diane S.
TITLE Measuring Writing Proficiency: A Registered Holistic Scoring Plan.
PUB DATE Apr 84
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Diagnosis; *Equated Scores; Grade 9; Graduation Requirements; High Schools; *Holistic Evaluation; Measurement Techniques; Rating Scales; *Scaling; Scoring; Standards; *State Programs; Student Evaluation; Testing Programs; Writing (Composition); *Writing Evaluation; Writing Skills
IDENTIFIERS New Jersey Department of Education; *Registered Holistic Scoring

ABSTRACT

The current New Jersey high school graduation law requires that students pass a ninth grade test in reading, writing, and mathematics. Equating the writing test from year to year to be sure that passing scores remain identical, regardless of the test form administered, requires developing a scoring technique. A research study was developed based on examining scoring consistency and maintaining the quantity and quality of scorer observations on any single essay response. Options for equating included: (1) using the current holistic scoring method and increasing the number of readers who rate each essay; (2) modifying the scoring procedure; and (3) adopting another scoring method. Methods offered for satisfying the constraints of equating were the Registered Holistic Score Scale, an analytic application to the Registered Holistic Score Scale and a variation of the Registered Holistic Score Scale divided into content and editorial features. The Registered Holistic Score Scale and corresponding Registered-Analytic Application allow for scoring consistency because of "registered" or established criteria. In addition, these consistent variables are invaluable in providing diagnostic information about a student's specific strengths or weaknesses in writing. Either process allows reporting of student performance on a comprehensive range of proficiencies considered important to good writing independent of the essay topic developed each year. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED270488

MEASURING WRITING PROFICIENCY

A REGISTERED HOLISTIC SCORING PLAN

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. S. Braungart-Bloom

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

By

Diane S. Braungart-Bloom
Bureau of Academic/Cognitive Skills
New Jersey State Department of Education
225 West State Street
Trenton, New Jersey 08625

ABSTRACT

In New Jersey the current high school graduation law requires students to pass a ninth grade statewide test in reading, writing and mathematics as one of the requirements for a high school diploma. Because new forms of the tests are developed annually, scores on each examination must be equated or anchored from year to year to make certain that the passing score remains identical regardless of the test form administered. Equating the reading and mathematics multiple choice tests is a relatively straightforward procedure, however, the writing test with its multiple choice section and single generic prompt is another issue.

Holistic scoring, as it was developed and intended by the Educational Testing Service (ETS), is a scoring method that uses a score scale that is relative to the essays scored in a given year. Therefore, in one year a "6" (the best writing identified within a given population of essays) might be a "4" relative to the following year's essay population. Given the nature of holistic scoring, and the fact that the essay topic changes from year to year (an uncontrolled variable), scorers respond to each paper as a "whole" without considering individual components such as organization, word use, style, grammer, etc. Scorers under this basic premise determine a paper's quality based upon its "total impression and the scores are distributed according to a normal distribution. Thus, recognizable and verifiable qualities of communication independent of topic and mode must be established to provide equivalent scoring from year to year. The concern lies in the fact that these "unknown" qualities may change because educator's instructional goals may have shifted from one year to the next (Eldridge, 1981).

Eldridge, Richard. "Grading in the 70s: How We Changed." College English, 43 (1981), 64-68

April 23-27, 1984

Presented at the 1984 Annual Meeting of the American Educational Research Association in New Orleans

TM 860 371

MEASURING WRITING PROFICIENCY

A REGISTERED HOLISTIC SCORING PLAN

Diane S. Braungart

New Jersey State Department of Education

Paper Presented at the

Annual Meeting of the National Council

on Measurement in Education

New Orleans, Louisianna

April 23-27, 1984

MEASURING WRITING PROFICIENCY

A REGISTERED HOLISTIC SCORING PLAN

Diane S. Braungart, Project Director

New Jersey State Department of Education

INTRODUCTION:

In New Jersey the current high school graduation law requires students to pass a ninth grade statewide test in reading, writing and mathematics as one of the requirements for a high school diploma. Until now, that requirement has been satisfied based on a minimum competency test in reading and mathematics. Beginning with the ninth grade class of 1985-86, students will be required to pass a new, more rigorous statewide test in reading, writing and mathematics as part of their high school graduation requirement.

Because new forms of the tests are developed annually, scores on each examination must be equated or anchored from year to year to make certain that the passing score remains identical regardless of the test form administered. Equating the reading and mathematics multiple choice tests is a relatively straightforward procedure, however, the writing test with its multiple choice section and single generic prompt is another issue.

Holistic scoring, as it was developed and intended by the Educational Testing Service (ETS), is a scoring method that uses a score scale that is relative to the essays scored in a given year. Therefore, in one year a "6" (the best writing identified within a given population of essays) might be a "4" relative to the following year's essay population. Given the nature of holistic scoring, and the fact that the essay topic changes from year to year (an uncontrolled variable), scorers respond to each paper as a "whole" without considering individual components such as organization, word use, style, grammar, etc. Scorers under this basic premise determine a paper's quality based upon its "total" impression and the scores are distributed according to a normal distribution. Thus, recognizable and verifiable qualities of communication independent of topic and mode must be established to provide equivalent scoring from year to year. The concern lies in the fact that these "unknown" qualities may change because educator's instructional goals may have shifted from one year to the next (Eldridge, 1981).

Eldridge, Richard. "Grading in the 70s: How We Changed." College English, 43 (1981), 64-68.

Since equating requires common items or common examiners -- with the holistic system we have neither -- the issue of equating essays becomes one of anchoring consistent variables from year to year. Therefore, given a statewide test that occurs annually with a change in topic and student population, consistency is not easily obtained. In an attempt to resolve this issue, we have developed a research study based on examining scoring consistency and maintaining the quantity and quality of scorer observations on any single essay response.

CONSIDERATIONS AND ALTERNATIVES

Options for equating include using the current holistic scoring method and increasing the number of readers who rate each essay, modifying the scoring procedure, or adopting another scoring method. Since the cost factor is significant when increasing the number of raters using the current holistic method, the two other options appear to be more viable alternatives.

Option 1

If the decision is to modify the existing holistic scoring procedure, then it must be done within the constraints of topic variability, scoring consistency, and quantity/quality of observations. An approach which we are trying in New Jersey, the "registered" holistic scoring method, could possibly satisfy these constraints.

Registered Holistic Scoring

Registered Holistic Scoring is a hybrid of pure "general impression" holistic scoring. In Registerd Holistic Scoring each score point reflects writing qualities determined by the Chief Readers at the conclusion of the first 1982-1983 Statewide Holistic Scoring Session. These "registered" qualities are considered characteristic of those papers assigned a given score point by the 220 readers. More important, these registered qualities were determined after all essays and had been holistically read and assigned two independent scores. (Qualities were determined following the June 27-30, 1983 Holistic Scoring Session.) The "registered" holistic method, therefore, makes explicit the internalized essay qualities that contribute to a reader's general impression and influence the reader's judgment about the total effect of the composition.

Registered Holistic Scoring is closely related to holistic scoring in that the assigned score (6 to 1) reflects a reader's general impression of each composition's feature quality overall. This registered scoring process differs from pure holistic scoring in three important ways. First, essay characteristics are determined before a scoring session is conducted. These essay features registered by the chief readers at the conclusion of the first year of holistic scoring (1982-1983) are considered to be inherent aspects of essays at a given score point. Second, the registered holistic process requires that readers focus their attention on a set of prescribed composition characteristics and assign a score (6 to 1) based upon the writer's level of facility with these features. Third, this process requires that each possible

score represents a specific level of proficiency in the complete set of composition characteristics being observed. Each point on the registered holistic score scale is "anchored" by statements that describe the proficiency level that the point represents. The registered holistic score, then, has particular instructional value because it is associated with a description of writing performance established before the set of essays was rated. In contrast, holistic scoring is based upon the quality of the piece of writing when judged on its overall success as a whole presentation without pre-referenced criteria and therefore cannot address diagnostic information about a students's specific writing strengths or weaknesses.

In summary, then, Registered Holistic Scoring (a modification of the holistic method) satisfies the constraints of equating in the following ways. First, the problem of topic variability is minimized since the predetermined criteria are applicable regardless of the mode of discourse. The criteria are not narrative or expository based; rather they are communication based. Second, scoring consistency is maintained because the statements within each category, at each score point, is based upon the following substantive information: 1) essay features or qualities registered by the confirmation of two chief readers following the holistic reading of 95,000 essays in 1982-83, and 2) resubstantiation of those qualities or features in 1983-84 by 240 readers who read holistically another set of 90,000 ninth grade student essays. (Resubstantiation will be attempted after the scoring session by asking the readers to respond to a verification survey form.) In addition, the predetermined qualities assigned to each score point (6 to 1) are used

in subsequent range-finder and scorer training sessions. Thus, scoring consistency is maintained and the criteria do not change regardless of the topic developed and used each year. Third, with "registered" holistic scoring, quantity of five features is retained because each score point has five statements related to five categories which represent the quality of communication inherent in essays assigned at that given score point. Those qualities are based upon elements considered necessary for students to communicate ideas to others through writing. Those elements or properties are found across five categories - organization, ideas, audience, syntactic structure and rhetorical modes, and mechanics, usage, sentence structure, and diction. (See Appendix A) When the categories are applied to any essay they are done so collectively at each score point. That is, all five feature statements must be exemplified in the essay assigned a specific score point. In other words, the scorers use "registered" or predetermined standards that are maintained and reinforced at each score point of the 6 to 1 score scale throughout the readings. Such predetermined, collectively verified standards afford students both consistency and quality in the readers' application of the score scale, from year to year--necessary requisites for equating. It is possible, therefore, in any given year to find there are no papers that represent a given score point because of the "registered" criteria within each category.

Experimental Study

In May, 1984 the New Jersey State Department of Education's Bureau of Educational Assessment will conduct a four day Registered Holistic Scoring Session using a sample of the 95,000 ninth grade student essays. The study

will be conducted by Measurement Incorporated of Durham, North Carolina - a subcontractor of National Computer Systems (test contractor). Measurement Incorporated will employ experienced readers for two separate studies. One group will be trained to apply the Registered Holistic Score scale and the other group to apply Option 2, that is, the Analytic Application of the Registered Holistic Method. A description of Option 2 follows.

Option 2

The second viable alternative is adopting or modifying another scoring method. One such approach could be to modify the Registered Holistic Score Scale through an analytic application. For example, student essays would be evaluated by readers within and across the five categories stated earlier. (See Appendix B, chart entitled Registered Holistic Score Scale: An Analytic Application.) Using Registered Holistic with an analytic application the reader is trained in applying the 6 to 1 score scale and corresponding statements for each of the five categories listed in the column on the right. This scoring method would yield a subscore in each area of the five categories along with a total score with respect to the writer's command of written English.

One advantage of this scoring method, in addition to the previously mentioned equating characteristics of consistency and quality, is that increased observations can be obtained without increasing the number of scorers; however, the training process and scoring sessions may be more lengthy. Applying the five category scale and 6 to 1 scores, it could be possible for a student to receive a maximum of 30 points - 5 categories x 6 (strong command); a secure number of observations in which to anchor or equate essays from one year to the next. In addition, the scale provides the student with diagnostic information across five proficiencies because the subscores focus on the components of writing and allow the equating factor to be dispersed over the five categories with a range of 6 to 1 score

points. Two independent readings of a given essay would double the score values and verify the score of the two readers. In contrast, the "registered" holistic score would be a collective score for all five categories doubled by a second independendent reading and the total score would reflect that the student's writing sample exhibited the qualities or features stated across the five categories. A point to keep in mind with the registered analytic approach is that since the method has the reader look at each aspect of the essay separately, the training process and scoring sessions may be longer.

Option 2: A Variation

A variation of the Registered Holistic Scoring Plan could be to divide the 5 categories according to a two part scheme that would measure writer proficiency in terms of CONTENT: 1) organization, ideas, and audience, and in terms of EDITORIAL aspects: 2) syntactic structure, rhetorical modes, mechanics, usage, sentence structure and diction. In this variation students would receive a holistic score of 6 to 1 in addition to scores for content and editorial aspects. The rater could assign a "+" or "-" to an essay based upon observed strengths (+) or weaknesses (-) in these two categories. An alternative method would be to assign a "C" if there are content problems observed or an "E" if the essay has problems related to editorial functions. Either approach to this variation could be applied quickly and provide diagnostic information (though limited) to teachers and parents. Equating would be addressed since the CONTENT and EDITORIAL features are consistent (related to qualities registered in 1982-83 and verified in 1983-84) and emphasize quantity as well as quality of the scorers' observations. Quantity is achieved by the student essay receiving two observations (content and editorial) by two scorers - 2 x 2 = 4 observations per essay. Quality is maintained because of the relative ease in applying a (+/- or C/E) descriptive score when using the registered features believed to constitute good communication.

CONCLUSION

Given New Jersey's attempt to establish equating procedures on an examination that students are required to pass as one of the requirements

for receiving a state endorsed high school diploma, the following elements need to be highlighted with respect to the Registered Holistic Score Scale and corresponding Registered-Analytic Application. First, scoring consistency is obtained because of "registered" or established criteria that was collected following the first (1982-83) and second (1983-84) year of holistic scoring of 95,000 students' essays respectively. These criteria are based upon properties considered necessary in order for students to communicate ideas to others through writing. Furthermore, these criteria are registered by chief readers and scorers alike following the two subsequent years of holistic scoring. Thus, beginning in 1984-85 the criteria becomes the "pre-determined" standard of quality and "typical" of papers at the given 1 to 6 score levels. In addition, the assigned scores registered in the first year by two chief readers and agreed upon by the readers in the second year, reflects consistency when applying a scale that determines the quality of students' writing ability from one year to the next.

Another advantage is that the score scale allows for far more extensive reporting than that which is obtained on the holistic score scale. In holistic reporting, a score is relative, that is, the score reflects a degree of writing ability based upon the entire group of essays being scored at that time. Using the "registered" score scale, an assigned score represents qualities of writing <u>across</u> ("registered" holistic) or <u>within</u> ("registered" analytic) five categories 1) organization, 2) ideas, 3) audience, 4) syntactic structure and rhetorical modes, and 5) mechanics, usage, sentence structure, and diction. Such information is valuable to equating since it allows for the variables to be anchored from year to year

because they are consistent or registered. In addition, these consistent variables are invaluable to students, parents and educators in providing diagnostic information about a given student's specific strengths or weaknesses in writing. Either process allows reporting of student performance on a comprehensive range of proficiencies, that is, the qualities considered important to all good writing independent of the essay topic developed each year.

APPENDIX A

REGISTERED HOLISTIC SCORING SCALE*

In each category of the six-point scale, attributes of a writer's command at written language have been identified. These attributes relate the individual essay to its particular place in the overall scoring process and are presented below:

6 = The response indicates a STRONG COMMAND of written language.

A writing sample in this category does not have to be perfect, but the sample will:

- be well developed and organized -- unified with focus on the assigned topic.
- develop key ideas with appropriate and varied details.
- appropriately address the specified or intended audience (educators).
- demonstrate syntactic and verbal sophistication through an effective variety of sentence structure and/or rhetorical modes.
- have very few, if any, errors in mechanics, usage, sentence structure, and diction.

*Diane S. Braungart
New Jersey Department of Education
Bureau of Educational Assessment

5 = The response indicates a GENERALLY STRONG COMMAND of written language.

A writing sample in this category will generally:

- be well developed and organized -- unified with focus on the assigned topic (minor flaws may be evident).

- develop key ideas with appropriate and varied details.

- appropriately address the specified or intended audience demonstrating at least an implicit awareness of the audience.

- demonstrate an awareness of syntactic and verbal sophistication through an effective variety of sentence structure and/or rhetorical modes.

- have few errors in mechanics, usage, sentence structure, and diction.

4 - The response indicates AN ADEQUATE COMMAND of written language.

A writing sample in this category will have an adequate command of the language and may be relatively bland, but the sample will:

- be adequately developed and organized on its own terms with unity and focus evident, but with some flaws.

- develop some of its key ideas with some appropriate and varied details.

- reflect an adequate sense of the intended audience by a usually consistent use of appropriate language.

- demonstrate a generally correct sense of syntax, with the avoidance of excessively monotonous pattern of sentence structure and/or rhetorical modes.

- display some errors in mechanics, usage, sentence structure, and diction but these will not substantially detract from the meaning of the response nor constitute a consistent pattern in any one of the cited categories.

3 = The response indicates a PARTIAL COMMAND of written language.

A writing sample in this category does not consistently demonstrate a command of the language and the sample will:

- be written on the assigned topic and be somewhat unified and focused. It may not have a beginning or an end; there may be difficulty in moving from idea to idea.

- to develop key ideas; unclear at times with few appropriate and/or varied details.

- reveal the writer's unawareness of an intended audience and/or address the specified audience inappropriately.

- demonstrate an excessive syntactic monotony in sentence structure and/or rhetorical modes.

- display a pattern of errors in mechanics, usage, sentence structure, or diction.

2 = The response indicates a LIMITED COMMAND of written language.

A writing sample in this category is generally inadequate and the sample will:

- display weaknesses in development and organization. It is somewhat unified and focused; it may not have a beginning or an end and the reader may encounter difficulty in moving from idea to idea.

- fail to develop key ideas; no elaboration; few, if any, appropriate details.

- reveal the writer's unawareness of an intended audience and address the audience inappropriately.

- demonstrate excessive syntactic monotony in sentence structure and/or rhetorical modes.

- display errors in mechanics, usage, sentence structure, or diction so severe to detract from the meaning of the response.

1 = The response indicates an INADEQUATE COMMAND of written language.

A writing sample in this category is manifestly inadequate and the sample will:

- relate to the topic, but be disorganized and lack a focus. (The reader finds it consistently difficult to follow.)

- fail to develop and/or state key ideas; details -- if any -- are inappropriate.

- reveal the writer's unawareness of any audience and if any audience is addressed, it is not the specified one.

- demonstrate an assortment of incorrect sentences and/or sentence fragments. Statements may be either incoherent or unintelligible.

- display numerous serious errors in mechanics, usage, sentence structure, or diction; reducing the response to an essay that is either incoherent or unintelligible.

APPENDIX B
REGISTERED HOLISTIC SCORE SCALE:
AN ANALYTIC APPLICATION

THE WRITER'S COMMAND OF WRITTEN ENGLISH IS ...	1 POOR	2 INADEQUATE	3 LIMITED	4 ADEQUATE	5 GOOD	6 EXCELLENT	ESSAY FEATURES
W	relates to the topic, but is disorganized and lacks a focus. (The reader finds it consistently difficult to follow.)	displays weaknesses in development and organization. It is somewhat unified and focused; it may not have a beginning or an end and the reader may encounter difficulty in moving from idea to idea.	is written on the assigned topic and is somewhat unified and focused. It may not have a beginning or an end; there may be difficulty in moving from idea to idea	is adequately developed and organized on its own terms with unity and focus evident, but with some flaws	is well developed and organized--unified with focus on the assigned topic (minor flaws may be evident).	is well developed and organized--unified with focus on the assigned topic.	ORGANIZATION
H E N	fails to develop and/or state key ideas; details--if any--are inappropriate.	fails to develop key ideas; no elaboration; few, if any, appropriate details.	fails to develop key ideas; unclear at times with few appropriate and/or varied details.	develops some of its key ideas with some appropriate and varied details.	develops key ideas with appropriate and varied details.	develops key ideas with appropriate and varied details.	IDEAS
T H E	reveals the writer's unawareness of any audience and if any audience is addressed, it is not the specified one.	reveals the writer's unawareness of an intended audience and addresses the audience inappropriately.	reveals the writer's unawareness of an intended audience and/or addresses the specified audience inappropriately.	reflects an adequate sense of the intended audience by a usually consistent use of appropriate language.	appropriately addresses the specified or intended audience demonstrating at least an implicit awareness.	appropriately addresses the specified or intended audience (educators).	AUDIENCE
E S S A Y	demonstrates an assortment of incorrect sentences and/or sentence fragments. Statements may be either incoherent or unintelligible.	demonstrates excessive syntactical monotony in sentence structure and/or rhetorical modes.	demonstrates an excessive syntactical monotony in sentence structure and/or rhetorical modes.	demonstrates a generally correct sense of syntax, with the avoidance of excessively monotonous pattern of sentence strucutre and/or rhetorical modes.	demonstrates an awareness of syntactic and verbal sophistication through an effective variety of sentence structure and/or rhetorical modes.	demonstrates syntactic and verbal sophistication through an effective variety of sentence structure and/or rhetorical modes.	SYNTACTICAL STRUCTURE RHETORICAL MODES
. . .	displays numerous serious errors in mechanics, usage, sentence structure, or diction; reducing the response to an essay that is either incoherent or unintelligible.	displays errors in mechanics, usage, sentence structure, or diction so severe to detract from the meaning of the response.	displays a pattern of errors in mechanics, usage, sentence structure or diction.	displays some errors in mechanics, usage, sentence structure, and diction but it does not substantially detract from the meaning of the response and does not constitute a consistent pattern in any one of these categories.	has few errors in mechanics, usage, sentence structure, and diction.	has very few, if any errors in mechanics, usage, sentence structure, and diction.	MECHANICS, USAGE, SENTENCE STRUCTURE DICTION

Diane S. Braungart
N.J. State Department of Education
Bureau of Educational Assessment

BEST COPY AVAILABLE